# SILICON PROCESSING

# FOR

# THE VLSI ERA

## VOLUME 2:

## PROCESS INTEGRATION

STANLEY WOLF Ph.D.
Professor, Department of Electrical Engineering
California State University, Long Beach
Long Beach, California

LATTICE PRESS

Sunset Beach, California

DISCLAIMER

This publication is based on sources and information believed to be reliable, but the authors and Lattice Press disclaim any warranty or liability based on or relating to the contents of this publication.

Published by:

Lattice Press,
Post Office Box 340
Sunset Beach, California 90742, U.S.A.

Cover design by Roy Montibon, Visionary Art Resources, Inc., Santa Ana, CA.



Library of Congress Cataloging in Publication Data
Wolf, Stanley

Silicon Processing for the VLSI Era
Volume 1 : Process Integration

Includes Index
1. Integrated circuits-Very large scale integration. 2. Silicon. I. Title

86-081923

ISBN 0-961672-4-5

9 8 7 6

PRINTED IN THE UNITED STATES OF AMERICA

to dope the polysilicon to a sheet resistance of 20-30 Ω/sq. This resistance is adequate for MOS circuits with gate lengths ≥3 μm. For smaller devices, polycide layers (i.e., composite layers of refractory metal silicides and polysilicon) can be used to reduce the sheet resistance to ~1 Ω/sq (see Vol. 1, chap. 11). Using a polycide gives us the benefits of both silicon-gate and metal-gate technologies.

The gate structure and polysilicon interconnect structures are then patterned using *Mask #4* (Fig. 5-15f). Following exposure and development of the resist, the polysilicon film is etched (in current technology this is done by means of a dry-etch process). This is a critical etch step for several reasons. First, the channel length of the device depends on the gate length, because of the self-aligned nature of the silicon gate technology. Hence, the gate-length dimension must be precisely maintained across the entire wafer, and from wafer to wafer. Second, the profile of the etched poly gate structure should be vertical; this will prevent variation of channel lengths by the penetration of the ions of the thinner regions of the gate sidewalls during formation of the source/drain regions by ion implantation. Third, to achieve the above goals, an anisotropic polysilicon etch process must be employed. This type of process, however, requires overetching to remove the locally thicker regions of polysilicon that exist wherever it crosses steps on the wafer surface. During the overetch time, areas of the thin gate oxide are exposed to the etchants. Thus, it is necessary to use a polysilicon etch process that is highly selective with respect to $SiO_2$.

**5.4.1.5 Formation of the Source and Drain Regions.** Once the gate has been fabricated, the source and drain regions can be formed. This is normally done by ion implantation without the use of a lithography step (Fig. 5-15g). The gate and the field oxide act as masks to prevent the ion implantation from penetrating to the silicon substrate below. Therefore, only the active regions covered by the gate oxide (and no gate polysilicon), are implanted. An $n^+$ implant is used, with an energy that is insufficient to penetrate the gate-poly or field-oxide layers (arsenic is typically used, with a dose of ~$10^{16}$ atoms/cm$^2$ and an energy of 30-50 keV). As noted earlier, the source and drain are thereby *self-aligned* to the gate, and the dimension of the polysilicon gate thus plays a major role in the defining of the MOS gate length.

Following the source/drain implant, an anneal (or drive-in) step is performed to activate the implanted atoms and to position the source/drain junctions as desired. During this step, some of the phosphorus doping of the polysilicon outdiffuses into the silicon substrate wherever a buried contact opening the gate oxide has been cut. This diffusion (which occurs both vertically and laterally into the silicon below) forms a heavily doped $n^+$ region under the polysilicon in the buried-contact exposed region. The lateral diffusion of the implanted source/drain dopant thereby becomes electrically connected to the $n^+$ region under the polysilicon buried-layer region. In this manner, an electrical connection between the polysilicon and the silicon is established at the buried contact locations. In some processes, the junction formed by the buried-contact dopant outdiffusion from the polysilicon is deeper than the source/drain junctions, while in others it is not as deep.

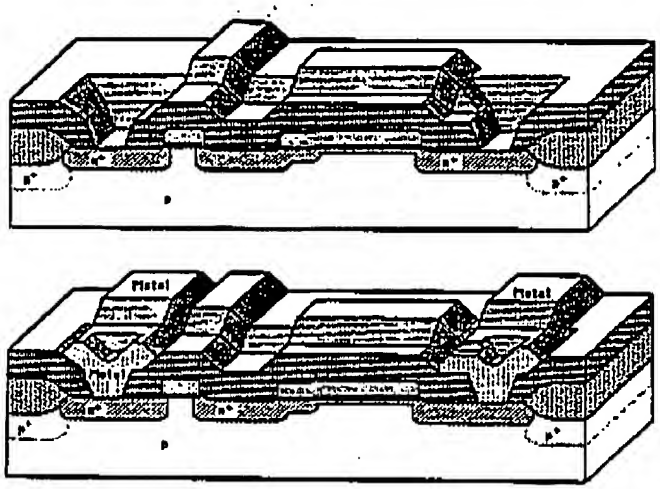


Fig. 5-15 (h) Contact cuts of the E-D inverter. Metallization of the E-D inverter. (i) Completed NMOS E-D inverter structure. Figures 15a through 15i from W. Maly, *Atlas of IC Technologies*. Copyright 1987 by the Benjamin/Cummings Publishing Company. Reprinted with permission.

The source/drain drive-in step also plays a part in determining the effective channel length ($L_{eff}$). That is, if the lateral junction depth is $x_{jl}$ (which is primarily determined by the the lateral diffusion during the drive-in step, because the lateral straggle of arsenic at 30 keV is only ~5 nm, see Vol. 1, chap. 9), $L_{eff}$ will be decreased by $2x_{jl}$ from the gate length at the mask level. Note that the channel *width* is also reduced by the bird's beak encroachment into the active area (see chap. 2). Thus, the actual width, $W_w$, of an MOS device is $W_w = W - \Delta W$, where $W$ is the width at the mask level and $\Delta W$ is the channel-width shrinkage during processing.

The depth of the source and drain is thus a critical dimension, but the doping concentration is not as important. (A discussion of shallow source/drain junction formation techniques is presented in chap. 3, section 3.10.) To a first approximation, the device characteristics will not depend on the doping concentration value, provided it is sufficiently heavy.

A diffusion step may be used to dope the source/drain regions. In some of these cases the dopant source of the diffusion is the CVD oxide layer that is deposited after the gate has been defined (see next section).

**5.4.1.6 Contact Formation.** After the source/drain regions have been formed, a CVD process is used to deposit a layer of doped $SiO_2$ (glass), about 1 μm thick, onto the wafers (see Vol. 1, chap. 6). The dopant in the $SiO_2$ is either phosphorus (in which

inversion layer at the two edges. Although this leads to a slightly higher $V_T$, the effect is much less severe than that observed in devices with heavy channel-stop implants.

## 5.5.2 Short-Channel Effects on Subthreshold Currents (Punchthrough and Drain-Induced Barrier Lowering)

In section 5.2.3 we described the nature of subthreshold current flow ($I_{Dst}$) in MOSFETs, noting that a specific value of the *subthreshold-swing* parameter (S.S.) can be attributed to such "normal" $I_{Dst}$ currents in long-channel devices. In short-channel MOSFETs, however, larger $I_{Dst}$ values are observed at lower voltages than predicted by long-channel device models: one manifestation is an increase in the value of S.S. Note that even relatively small values of $I_{Dst}$ can limit the transistor's ability to isolate nodes in a dynamic circuit or can allow excess current in static inverters. Hence, care must be taken to minimize $I_{Dst}$. Two of the primary causes of increased $I_{Dst}$ are *punchthrough* and *drain-induced barrier lowering* (DIBL).

Punchthrough is normally observed when the gate voltage is well below $V_T$. It occurs as a result of the widening of the drain depletion region when the reverse-bias voltage on the drain is increased. The electric field of the drain may eventually penetrate into the source region and thereby reduce the potential energy barrier of the source-to-body junction (Fig. 5-19).[23] When this occurs, more majority carriers in the source region have enough energy to overcome the barrier, and an increased current then flows from source to body. Some of this current is collected by the drain, thereby increasing $I_{Dst}$. In general, punchthrough current begins to dominate $I_{Dst}$ when the drain and

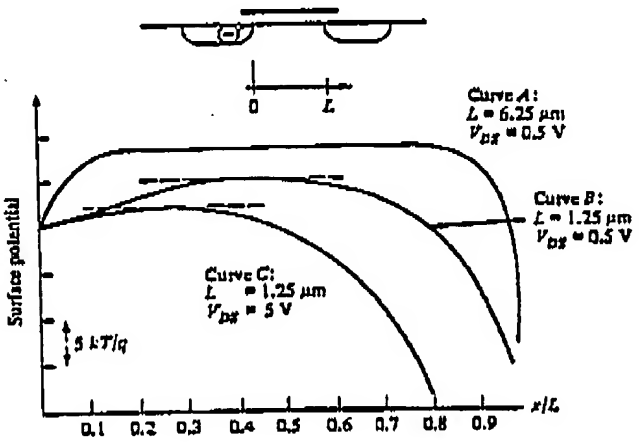


Fig. 5-19 Surface potential in the channel for devices with different channel lengths.[23] (© 1979 IEEE).

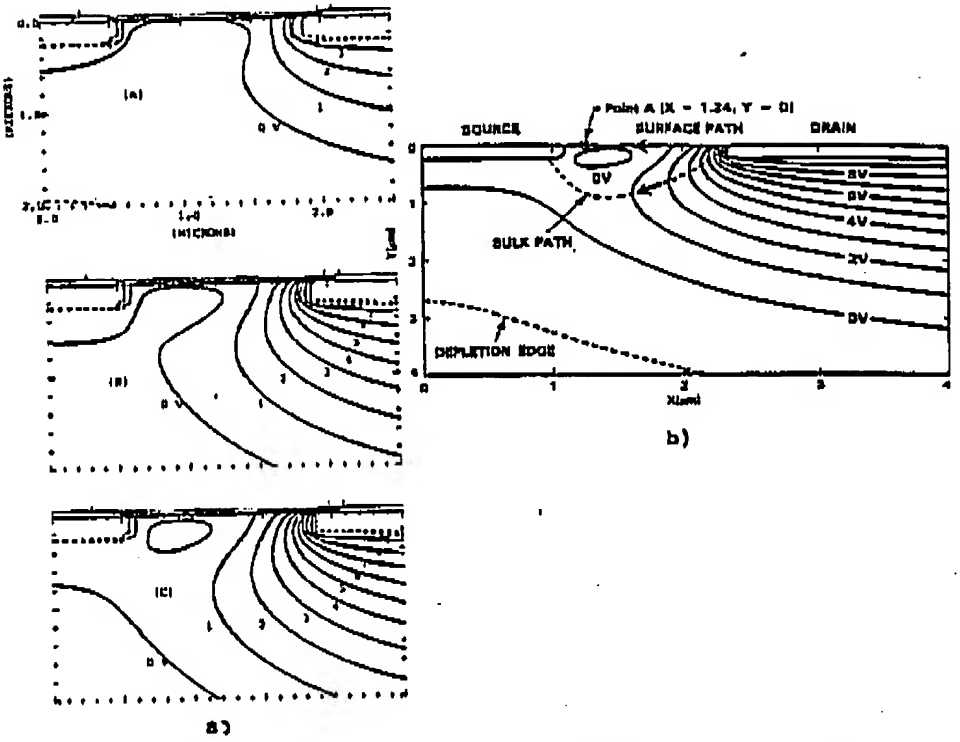


Fig. 5-20 (a) 2-D potential profile of an *n*-channel MOSFET with a drain bias of: 1) 3 V; 2) 7 V; 3) 9 V. Channel length = 1 $\mu$m. (b) Simulation of the potential profile of an *n*-channel MOSFET with a gate and drain bias of 0 and 9 V respectively. The surface DIBL and bulk punchthrough paths are indicated. From K. M. Cham, *et al.*, *Computer Aided Design and VLSI Device Development*. Copyright 1986 Kluwer Academic Publishers. Reprinted with permission.

source depletion regions meet, and it can be suppressed by keeping the total width of the two depletion regions smaller than the channel length.[24]

Calculations of the potential in the bulk channel region in devices that use ion implantation to adjust $V_T$ indicate that the barrier is lowest away from the Si-$SiO_2$ interface (usually at almost the same depth as the source/drain junction depths). That is, the $V_T$-adjust implant increases the doping concentration near the surface of the channel, causing the drain depletion region to be wider in the bulk than it is near the Si-$SiO_2$ interface. As a result, punchthrough current flows *below* the surface (Fig. 5-20). Consequently, the gate voltage has less control over the subthreshold current (i.e., even with sufficient gate voltage to turn off the channel, $I_{Dst}$ can still flow in such devices).

An enhancement-mode device which is not turned off when $V_G = 0$ loses its ability to function as a switch.

Similarly, the application of a drain voltage in short-channel devices can also cause drain-induced barrier lowering (DIBL). That is, the drain voltage can cause the *surface potential* to be lowered (Fig. 5-21).[23, 71] As a result, the potential energy barrier at the *surface* will be lowered, and the subthreshold current in the channel region at the Si-$SiO_2$ interface can be increased (5-21b). This implies that $I_{Dst}$ at the surface due to DIBL is expected to become larger as the gate voltage approaches $V_T$.

These two effects illustrate the complexity involved in modeling the overall subthreshold I-V behavior of short-channel MOSFETs. That is, both punchthrough current (in the bulk), as well as DIBL-induced current (at the surface), may simultaneously contribute to $I_{Dst}$.

To prevent punchthrough current in short-channel devices, the substrate doping can be increased to decrease the depletion-layer widths. These widths can be estimated using the formula for the width of a one-sided step junction:

$$W = \sqrt{\frac{2 K_s \varepsilon_0 (|V_A| + V_{bi})}{q N_B}} \qquad (5-17)$$

where the built-in voltage, $V_{bi}$, given by

$$V_{bi} = 0.56 + (kT/q) \ln (N_B/n_i) \qquad (5-18)$$

and where $V_A$ is the total applied voltage and $N_B$ is the doping concentration of the body. Figure 5-22 gives the depletion-layer width of *pn* junctions as a function of doping and applied voltage.

However, increasing the substrate doping also increases the source-to-body and drain-

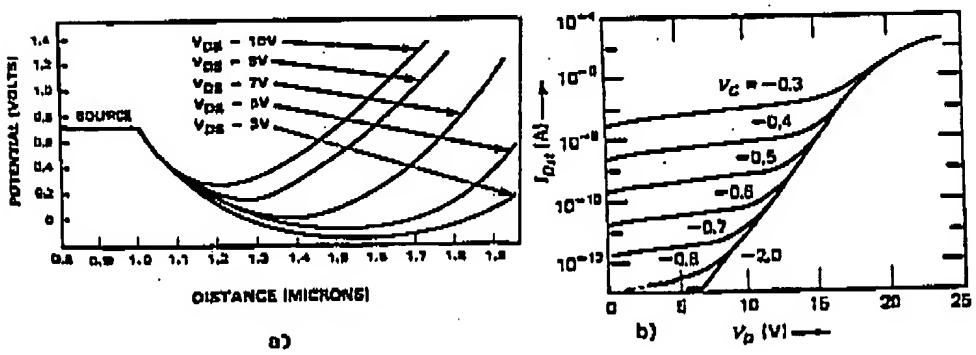


Fig. 5-21 (a) DIBL versus drain bias for short-channel MOSFET. (b) Experimental low-current characteristics for a MOSFET with $L = 2.1\ \mu m$, $V_{SB} = 0$.[81] (© 1974 IEEE).

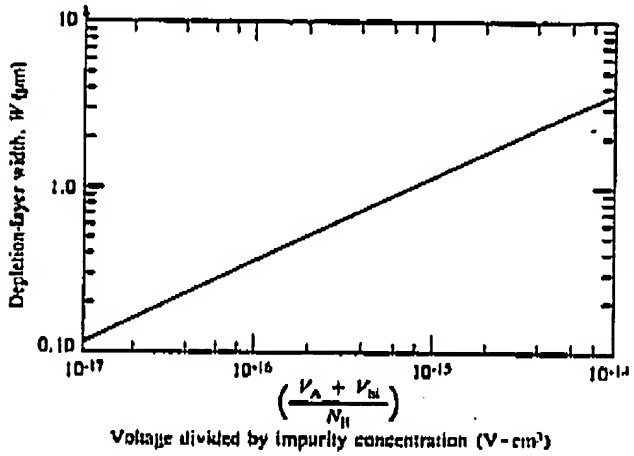


Fig. 5-22 Depletion-layer width of a one-sided step function as a function of doping and applied voltage calculated from Eqs. 5-17 and 5-18.

to-body junction capacitances, as well as the body factor. In addition, it reduces the breakdown voltages of the source/drain junctions. To avoid these drawbacks, an additional boron implant (whose peak concentration is located at a depth near the bottom of the source-drain regions) can be performed. This additional doping reduces the lateral widening of the drain-depletion region below the surface without increasing the doping under the junction regions. With such implants, the component of the punchthrough current can be suppressed to well below the normal $I_{Dst}$ current of the device.

For example, in Fig. 5-23, a 1.2-$\mu$m device with a body doping of $1.9 \times 10^{15}$ cm$^{-3}$ without such a punchthrough-stopping implant shows a large value of $I_{Dst}$ even when $V_G = 0$ V (curve A). This indicates that the device is already exhibiting punchthrough.[26] Implants of boron with a dose of $8 \times 10^{11}$ atoms/cm$^2$ and different energies are then performed in an attempt to reduce $I_{Dst}$ to the values exhibited by a long-channel device (curve B). If the implant is too shallow, the extra implant has the effect of shifting the $V_T$ of the device to well beyond the desired value. When the energy is increased so that the implant is sufficiently deep, the value of $I_{Dst}$ drops to that exhibited by the long-channel device. At the same time, the surface concentration remains essentially unchanged, so that $V_T$ is not appreciably shifted.

In another example, the S.S. of a device without a punchthrough-prevention implant is measured as its length is varied (Fig. 5-24a). At an $L_{eff}$ of ~0.85 $\mu$m the S.S. starts to increase, indicating that punchthrough current begins to dominate $I_{Dst}$. By adding an implant step that places boron atoms in the dashed subsurface region shown in Fig. 5-24b, the punchthrough component of $I_{Dst}$ is suppressed so that it is not observed until $L_{eff}$ becomes nearly as small as 0.5 $\mu$m.

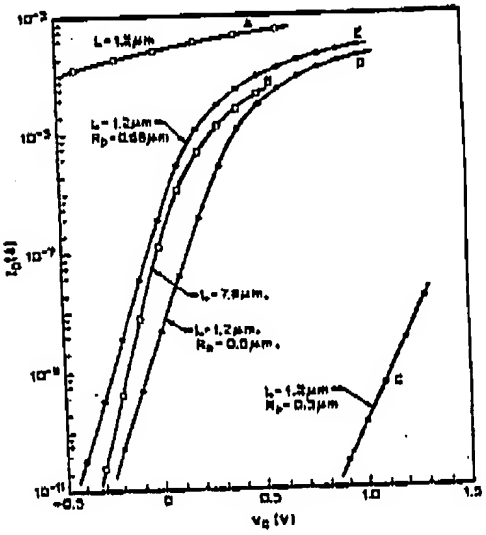


Fig. 5-23 Drain current versus gate voltage for *n*-channel devices with a substrate doping of $1.9 \times 10^{15}$ atoms/cm$^3$, source/drain junctions 0.47-μm deep, 575 Å gate oxide, drain voltage of 5 V, and $V_{BS}$ of 0 V. Devices A and B have no channel implant, and devices C and E have a boron channel implant of $8 \times 10^{11}$ atoms/cm$^2$ at various energies.[26] (© 1978 IEEE).

## 5.5.3 Short-Channel Effects on I-V Characteristics

The I-V characteristics of short-channel devices are significantly altered in three ways. First, the combined effects of reduced gate length and gate width produce a change in $V_T$. Second, the channel length is modulated by the drain voltage when the device is in saturation (i.e., $V_{DS} > [V_G - V_T]$), causing an increase in device gain over that predicted in an ideal long-channel device (*channel-modulation effect*). Third, the mobility of the carriers in the channel is reduced by two effects, which also reduces $I_D$. (The two effects are the *mobility-degradation factor*, due to the gate field, and the *velocity-saturation factor*).

Figure 5-25 shows the I-V characteristics of an MOS device.[28] The curves in Fig. 5-25a are those of an ideal long-channel MOS device, while those in Fig. 5-25b show the effect of adding the channel modulation factor. Figure 5-25c shows the combined effect of adding the mobility degradation factor to those of Fig. 5-25b. The velocity saturation factor also has the effect of making the both $I_D$ and $g_m$ independent of channel length in silicon MOS transistors for $L_{eff} \leq 1.25$ μm. More details on these effects are provided in suitable device physics texts.[1,3,28]

A general guide to the design of short-channel MOSFETs is given in references 62 and 69. In addition, a simple *engineering model* for short-channel devices has also been developed.[68] Its purpose is to provide a simple picture of the essential electrical behaviors of the short-channel MOSFETs from the perspective of a circuit designer. That is, this engineering model relates the terminal voltages to the drain current, much as Eqs. 5-8 and 5-10 yield the I-V characteristics for long-channel MOS devices. Consequently, device designers who need to relate device and process parameters to circuit parameters should also find this model useful.

## 5.5.4 Summary of Short-Channel Effects on the Fabrication of MOS ICs

In the first section of this chapter, we showed that the use of lightly doped substrates generally produced optimum device behavior in long-channel MOS transistors. In this

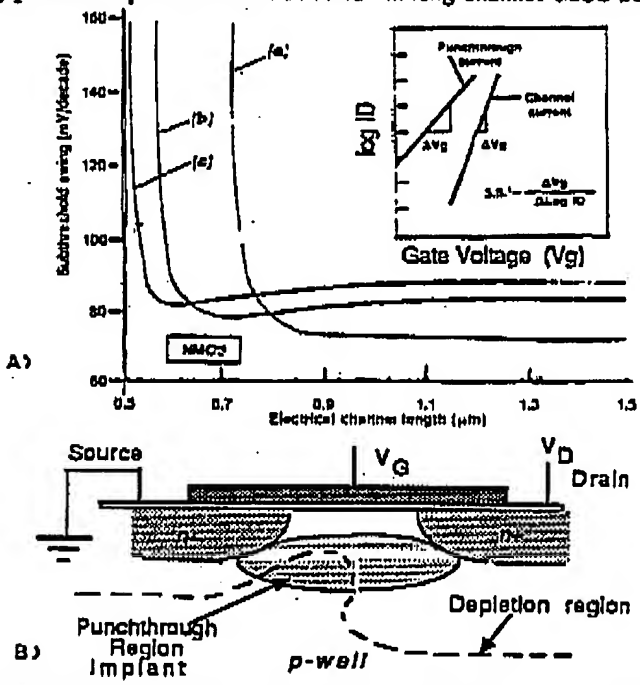


Fig. 5-24 (A) Subthreshold slope versus electrical channel length for NMOS devices ($V_{Tn}$ = = 0.7 V), having a common threshold adjustment implant and punchthrough implant doses of: of: (a) zero; (b) $2x10^{11}$ $cm^{-2}$; and (c) $3x10^{11}$ $cm^{-2}$.[27] Reprinted with permission of Semiconductor International. (B) NMOS cross-section with implant placed into punchthrough region.

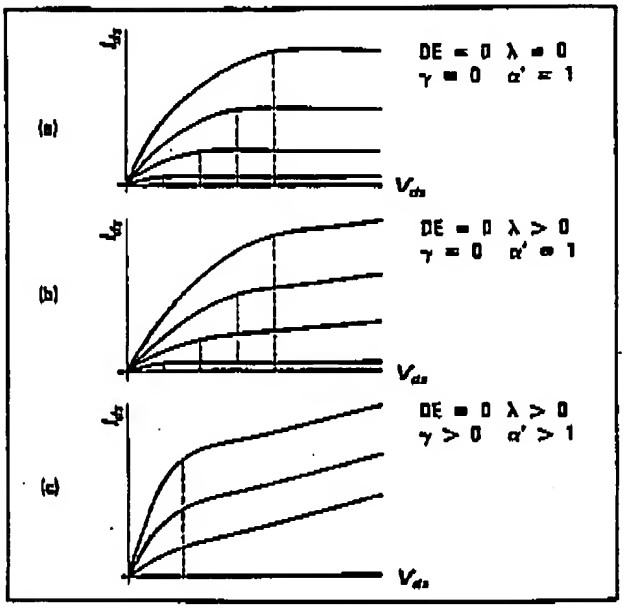


Fig. 5-25 The I-V curves of an MOS device showing the the effects of progressively increasing short-channel behavior. (a) Long-channel behavior. (b) With channel-length modulation. (c) Addition of velocity saturation.[28] (© 1986 IEEE).

section we noted that higher substrate doping is needed to overcome some of the detrimental impacts of short-channel effects. Thus, trade offs need to be made in selecting the proper substrate doping-concentration values to achieve optimum short-channel MOS device performance. Some of these trade offs are discussed by Kakumu,[29] who points out that a higher substrate doping concentration produces decreased ring-oscillator gate delay in submicron CMOS because of increased junction capacitances and decreased carrier mobility (due to increased impurity scattering).